DOCUMENT RESUME

ED 471 865 EA 032 099

AUTHOR	Hess, Frederick M.
TITLE	I Say "Refining," You Say "Retreating": The Politics of High-Stakes Accountability.
REPORT NO	PEPG/02-11
PUB DATE	2002-06-00
NOTE	61p.; Paper prepared for the "Taking Account of Accountability: Assessing Politics and Policy" Conference (Cambridge, MA, June 10-11, 2002).
AVAILABLE FROM	For full text: http://www.ksg.harvard.edu/pepg/TAConfPapers.htm.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Accountability; Administrator Attitudes; Educational Policy; Elementary Secondary Education; Evaluation Utilization; *High Stakes Tests; Political Attitudes; *Program Implementation; *Public Opinion; *Resistance to Change; School Policy

ABSTRACT

This paper examines ways in which context influences the implementation of high-stakes accountability systems. The trajectory of implementation efforts has followed a familiar course, with the emergence of scattered opposition prompting officials to "refine" testing systems in predictable ways. These efforts placate critics while softening the coercive impact of accountability. Whether these revisions eviscerate the larger system depends largely on the balance of political pressure. The paper addresses such questions as "Why do high-stakes accountability systems launched to widespread acclaim meet growing pockets of resistance even as student performance soars? Why are accountability provisions softened or made more flexible in predictable ways? and What are the implications of these issues for the promise of accountability-driven reform?" The paper outlines the general political dynamic by discussing the minimum-competency-testing push of 2 decades ago. It then surveys more recent efforts in California, Massachusetts, Texas, and Virginia to distill some insights regarding the role of context in the politics of coercive accountability. It argues that the fate of high-stakes reform turns on the willingness of the public and officials to accept high levels of concentrated costs and on the relative strength enjoyed by key critics. (Contains 101 endnotes.) (RJM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 471 865

I Say 'Refining,' You Say 'Retreating': The Politics of High-Stakes Accountability

Frederick M. Hess
University of Virginia

PEPG/02-11

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
F. M. Hess
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

fmh3x@virginia.edu

EA032049

Prepared for the Conference on "Taking Account of Accountability: Assessing Politics and Policy." Harvard Program on Education Policy and Governance. Cambridge, MA. June 10-11, 2002.

BEST COPY AVAILABLE

Abstract. Across the nation, high-stakes accountability systems adopted during the mid- and late-1990s are grinding towards implementation. The trajectory of these efforts follows a familiar course, with the emergence of scattered opposition prompting officials to "refine" testing systems in predictable ways. These efforts placate critics while softening the coercive impact of accountability. Whether these revisions eviscerate the larger system depends largely on the balance of political pressure. Here, I build on my earlier work in this area by generating concrete hypotheses about how context may influence this political dance. I introduce the general political dynamic by discussing the minimum competency testing push of two decades ago and then survey more recent efforts in California, Massachusetts, Texas, and Virginia in order to distill some insights regarding the role of context in the politics of coercive accountability. The fate of high-stakes reforms turns on the willingness of the public and officials to accept high levels of concentrated costs and on the relative strength enjoyed by key critics.

Introduction[*]

Across the nation, high-stakes accountability systems adopted during the mid- and late-1990s are grinding towards implementation. These efforts follow a familiar trajectory, with abysmal early student scores quickly improving even as scattered opposition begins to coalesce. When opposition reaches a certain level of intensity, officials seek to mollify critics by "refining" testing systems in ways purported to make them fairer and more rational. The challenge is for public officials to make such revisions without undermining essential elements of accountability.

Accountability efforts have occasioned extensive consideration of the merits of various tests, appropriate measurement techniques, and the design of these systems. Receiving far less attention have been the political tensions that shape—and frequently imperil—any push for high-stakes accountability, though these tensions prove to be as educationally significant as any technical components of accountability programs. In fact, surveying the developments and implementation of high-stakes accountability programs, it is no simple matter to determine which programmatic decisions are inspired by educational concerns and which are politically motivated. While high-stakes accountability is appealing in the abstract, implementation produces visible costs that are more politically salient, at least in the short term, than the educational benefits. The benefits are diffuse and long-term while the costs are immediate and concentrated, framing a struggle in which the complainants bear the upper hand.

The 1970s minimum competency movement was the first time this contest played out in the U.S., as a flood states adopted widely supported testing programs that called for

[*] The author would like to thank Amy Klekotka for her valuable research assistance.

students to master particular skills and content before graduating. In almost every case, large numbers of children failed to meet the initial standards but only an invisible handful of children were ever denied diplomas. While proponents hailed this pattern as a demonstration that testing had driven systemic improvement, a complementary development was a largely unheralded tale of political accommodation and compromise.

I do not mean to imply that the politics of accountability are static, across either time or place. In fact, a consideration of the past three decades suggests a growing social acceptance of substantive accountability. In particular, there appears to be a growing willingness on the part of voters and public officials to stand fast in the face of inequities and concentrated costs that sank earlier accountability efforts. Even against this backdrop, however, the politics of accountability have played out very differently in various states. This raises the possibility that the latest wave of state efforts will deliver the substantive changes that have often proved elusive.

Why do high-stakes accountability systems launched to widespread acclaim meet growing pockets of resistance even as student performance soars? Why are accountability provisions softened or made more flexible in predictable ways? How does state context help to explain these developments? Finally, what are the implications of these issues for the promise of accountability-driven reform?

The Politics of High-Stakes Accountability

The political challenges posed by accountability are a direct consequence of its educational promise. The allure of standards-based reform is straightforward. Standards are a statement that—at a minimum—schools ought to teach children certain knowledge and skills and that the state should ensure that children learn and schools teach this

content to a specified level of mastery. The resulting challenge is clear. Setting meaningful performance standards makes it inevitable that some students, teachers, and schools will fail to meet those standards. This poses a daunting political challenge in a democratic society where the "low-performers" have powerful incentives to challenge the legitimacy of the system.

It is important to distinguish between the "high-stakes" accountability systems that include sanctions for students and/or teachers and those nonintrusive systems that do not. High-stakes accountability systems link rewards and punishments to demonstrated student performance in an effort to transform the quality of schooling. Such systems press students to master specified content and force educators to effectively teach that content. Under such a regime, school improvement no longer rests primarily upon individual volition or intrinsic motivation. Instead, students and teachers are compelled to cooperate through levers such as diplomas and job security. Such "transformative" systems seek to harness the self-interest of students and educators in order to refocus schools and redefine the expectations of teachers and learners.[1]

These high-stakes efforts are fundamentally different from standards-based reforms that reject the coercive force of self-interest. Gentler, more suggestive standards-based approaches seek to improve schooling through informal social pressures, by using tests as a diagnostic device, by increasing coordination across schools and classrooms, and by using standardization to permit more efficient use of school resources. Suggestive accountability can produce educational benefits, but they tend to be modest and dependent on the ability and inclination of teachers to use the tests as pedagogical tools.

In practice, the two visions of standards represent two ends of a continuum. Many accountability programs begin with at least a rhetorical commitment to the transformative "high-stakes" ideal. Over time, however, implementation gradually reveals the costs implied by such change, eroding support for coercive accountability while opposition coalesces. In fact, opponents of transformative accountability only rarely suggest that they are opposed to the idea of accountability. Rather, they explain that their opposition is due to the nature of existing arrangements or to an exaggerated reliance on test results and insufficient attention to other measures of performance. Such critics implicitly agree that they will support transformative accountability if only… it is stripped of its transformative character.

In the face of such pressure, transformative systems are generally weakened and rendered more suggestive in one of five ways. Officials lower the stakes, make the test easier, reduce the thresholds required to pass, permit some students to side-step the required assessment, or delay the implementation of the exam. While each alteration is a response to legitimate programmatic concerns, the common thread is the manner in which they ease political resistance by weakening the coercive impact of accountability.

The Promise of Outcome Accountability

Conventionally, officials have judged public schools and school personnel on the basis of whether or not they comply with regulations and directives governing inputs, rather than upon student performance or progress. In large part, this approach was a compromise among policymakers unwilling or unable to resolve disputes regarding what schools should focus on or how school performance ought to be measured. Instead of pursuing an elusive consensus on such questions, officials accepted a "shopping mall"

ideal in which schools provided a smorgasbord of services.[2] This approach has resulted in well-documented problems, especially a lack of focus, uneven performance, and an effort to combat these problems through intrusive regulation and micro-management.

Consequently, there is broad support for the notion of "outcome accountability," in which states merely focus on establishing performance criteria and ensuring that they are met. Conceptually, outcome accountability offers a number of advantages. Specifying what skills and knowledge students are responsible for mastering fosters agreement on educational goals, giving educators clear direction. This enables administrators to more readily gauge teacher effectiveness and to respond by taking steps to mentor or motivate less effective teachers, and to recognize and reward effective teachers. Clear expectations and information on performance can ensure that "hard-to-educate" students are adequately served and make it difficult for schools to casually overlook such students or argue that they are being served "adequately." High-stakes accountability can enhance educator professionalism and boost public support for schooling by holding educators to clear standards and sanctioning those who do not meet them. Finally, an outcome focus can potentially boost pedagogical freedom by allowing supervisors to forego input regulation and concentrate on monitoring outcomes.[3]

Such changes may come at a price. High-stakes accountability may adversely alter the culture of schooling, narrow the scope of instruction and services that schools provide, leave less room for creative engagement, or shift educational resources into test-specific preparation.[4] Whether such changes outweigh potential benefits rests primarily on normative views of what schools are for and what constitutes good teaching.[5]

High-stakes accountability systems tend to show rapid student improvement in the initial years. In part, this is because the systems produce the desired behavioral changes among school officials, teachers, and students that improve student learning in the intended fashion. In part, it is also because schools adopt practices that boost student test performance without enhancing overall learning. Such responses can be constructive, if they direct attention to students or to skills that were previously being shortchanged, but they can also be distracting or counterproductive. Such distracting or counterproductive practices include increasing test preparation at the expense of substantive content or holding students back a year so that they will be better prepared for testing.

The Politics of Accountability

High-stakes accountability requires officials to make five politically sensitive sets of decisions. First, it is necessary to designate a prescribed body of content and objectives to be tested. Such a course necessarily marginalizes some of the goals, objectives, content, and skills that are not included. Second, it is necessary to impose assessments that render clear indications as to whether students have or have not mastered the requisite skills and content. Third, such assessment requires policymakers to specify what constitutes mastery. Fourth, designers need to decide what to do with students who fail to demonstrate mastery. Finally, if accountability is to significantly alter educational provision, educators must be rewarded or sanctioned on the basis of student performance. Each decision tends to produce passionate opposition among those who bear the costs of each choice.

Resisting the protests of the aggrieved is the central political challenge confronting advocates of high-stakes reform. In the face of heated opposition,

proponents often agree to a series of compromises on program design and implementation, eventually undercutting the promise of accountability. The result is that the primary benefits of accountability are usually the diagnostic benefits and the informal social pressures produced by publicizing test results, rather than the coercive promise implied by high-stakes testing.

The Few, the Angry, the Mobilized

Resistance to high-stakes reforms typically emerges among four constituencies: ethnic and socio-economic communities in which students are disproportionately sanctioned by tests, communities with well-regarded schools that resent the "disruption" or reputational threat of testing, educators concerned about their professional autonomy and the specter of sanctions, and those who find their moral or curricular preferences marginalized by the testing regime in question.

First, while accountability may yield significant long-term and systemic benefits regardless of individual losses or specific inequities, it does require penalizing some students. Of course, the current system takes a high toll on students who perform poorly, permitting many students to be promoted without mastering important skills or to graduate with meaningless diplomas.[6] The difference is that existing inequities can be attributed to impersonal social forces, while high-stakes accountability forces public officials to visibly sanction vulnerable children. Those students denied diplomas suffer clear and immediate costs, while the benefits of effective accountability tend to be diffuse and long-term. Those who lose out under high-stakes testing, because they have more immediately at stake, will tend to be passionate; the larger mass of "winners" will find the issue less pressing.[7] Even if disadvantaged children are the primary beneficiaries of

accountability systems—as many proponents argue—such benefits are indirect and hard to define. This situation is especially thorny because children in minority and low-income communities are disproportionately likely to fail high-stakes exams, leaving officials vulnerable to charges of callousness and racial bias. As a result, officials will find themselves pressured to reduce the number of failing students or to reduce the consequences of failure. For instance, for nearly a quarter century, the NAACP has officially opposed decisions to withhold diplomas or grade promotion on the basis of test results, deeming such policies an effort to blame the victim.[8]

Second, in the most highly regarded school systems there is also concern about the impact of high-stakes accountability. In these communities, the parents and educators are less concerned that students will be sanctioned than that an emphasis on state-mandated tests will hurt local schools by forcing them to shift their attention to state-dictated curricula and content. In particular, parents and educators in highly-regarded districts fear that the pressure to teach baseline skills and content will disrupt gifted, advanced placement, and International Baccalaureate classrooms. They also worry that the test scores are a inaccurate proxy for the broader quality of schooling, and that an emphasis on test scores may have a variety of negative consequences, such as understating school performance, impeding students' college prospects, and reducing local property values. While it can be readily argued that any disruptions indicate that all students were previously not mastering necessary skills and content, or that disappointing test scores may suggest that "elite" districts are not as effective as parents believe, there exists no natural constituency to advance these points. Meanwhile, the educated,

wealthy, and politically involved residents of high-performing suburban districts have a visceral desire to protect the practices and the reputations of their schools.

Third, teachers are generally averse to being evaluated or sanctioned on the basis of student performance.[9] It can be difficult for public officials to resist the concerted opposition of teachers, especially given the lack of a natural "pro-sanction" constituency. Meanwhile, children vary in ability and preparation from community to community and school to school, confronting some educators with greater challenges than others. This raises concern over whether officials can equitably determine teacher performance, forcing advocates of high-stakes accountability to defend inequities in the face of heated criticism from teachers and their allies.[10]

Teachers also have a second complaint, one more geared to the culture of schooling. For several decades, the American public education establishment has embraced a vision of professional, autonomous teachers who operate out of a sense of duty and commitment.[11] Whatever the strengths or weaknesses of such a system, it is the one to which current teachers have grown accustomed and in which they have been acculturated. The premise of high-stakes testing challenges the existing "schoolhouse" culture by pressing teachers to teach the content and skills mandated by the state, regardless of their personal preferences. In doing so, high-stakes testing also alters the low-pressure culture that educators have enjoyed. Educators have incentives to resist a system that challenges their autonomy, holds them accountable, and forces them to engage in practices they may not favor.

Finally, the multiple agendas that co-exist within public schooling ensure that the push for high-stakes accountability will provoke conflict from those whose particular

agendas will be marginalized. There is deep-rooted disagreement in the U.S. as to what schools are for, what a good education includes, and what skills and content children need to know. Efforts to impose statewide agreement will inevitably offend some constituencies.[12] The aggrieved will challenge the efforts to marginalize their concerns, while the broader public will have little cause for action.

In each case, proponents of standards must marshal diffuse support in response to challenges from aggrieved, passionate, coherent constituencies. The temptation for proponents is to compromise on the elements of accountability, slowly softening the coercive threat posed by high-stakes regimes. The American political system is notoriously bad at pursuing collective goods when it requires imposing concentrated costs on select groups. American government is highly permeable, making it relatively easy for small but passionate factions to block or soften adverse legislative or bureaucratic decisions.[13] The fate of accountability efforts often turns on the size, mobilization, and influence of these interests. In homogeneous states with weak teacher unions, for instance, the pressures on public officials to fundamentally compromise on coercive accountability is likely to be much less severe than in states with a strong union and large, influential minority communities. Generally, since visible and vocal groups like teachers and minorities are likely to feel like they are suffering large costs under high-stakes accountability, and given that the benefits of transformative accountability are generally diffuse and long-term, the political calculus favors the opponents of coercive accountability.

However, if an accountability system survives implementation, the political calculus may reverse. In time, accountability systems come increasingly to be seen as

central to legitimate schooling. They become intrinsic to the "grammar of schooling."[14] This is what long ago took place in Japan and Western Europe. Moreover, the existence of a widely accepted assessment regime can be useful to educators and public officials, permitting schools to concretely demonstrate performance and strengthen their claim on public support and resources.

Political pressures pinch more tightly as accountability shifts from concept to reality. The result is a cyclical dance, in which early support for standards is followed by gradually increasing pressure to lower the bar, shrink the number of "at-risk" parties, or lessen the consequences of failure. At its most basic, the politics of accountability is a desperate contest in which proponents race to institutionalize the regime before resistance leads officials to start dismantling it.

The dance of accountability is shaped by political context. High-stakes accountability systems face political challenges that are much stiffer in some states than in others. One object of this paper is to broaden my earlier work in this area by beginning to illuminate more clearly a few of the key contextual elements and how they may matter.[15]

Prologue: Minimum Competency Testing

The tensions that beset high-stakes accountability systems work are not new. They accompany any effort to institute high-stakes accountability. Perhaps the most relevant example prior to the current era was that of minimum competency testing in the 1970s, where a wave of states adopted exit tests that became a condition for receiving a high school diploma.

In the end, minimum competency testing proved a mixed success, but one unable to deliver on its ambitious promise. Few students were ever denied diplomas, in part because resources and tutoring were targeted to low-achieving students and more attention was paid to the academic preparation of previously overlooked students. In large part, however, passing rates were boosted by making the tests exceedingly simple, creating a low bar for passage, offering students a number of chances to pass, and exempting categories of low-performing students. In fact, the heavily compromised nature of the 1970s push for minimum competency testing would later help to engender support for more rigorous accountability measures in the 1990s.

After "minimum competency testing" was first introduced in Oregon in 1973, states began to adopt modified versions of the Oregon system. By 1979, spurred by concern that schools were no longer delivering essential instruction and that students were not mastering vital skills, thirty-six states had enacted some form of minimum competency testing.[16] Eighteen states required students to pass the tests for graduation, with almost all of these exclusively targeting reading, writing, and arithmetic. While the National Institute of Education observed that there was no uniform definition of minimum competency testing, such programs all sought to ensure that graduates mastered a small body of essential knowledge and skills.

Proponents effectively framed the question as one of whether or not states ought to demand some degree of educational performance. Because states had no way to ensure that students were mastering essential skills, and because it was easy to argue that literacy and numeracy were skills vital to any child's life chances, few opponents emerged and those that did made little headway. As one critic conceded in 1984,

"Promis[ing] a simple remedy for complicated problems of achievement and accountability, MCT has reached almost universal application in less than ten years."[17] By focusing on basic academic subjects where the public largely agreed about what graduates needed to know, proponents avoided messy debates about how to define essential knowledge or skills.

While minimum competency testing was typically enacted with only modest opposition, implementation would generate serious political and legal controversy. This was not at first apparent; MCT legislation normally stipulated that requirements would not apply to current high school students, creating a lag of at least four years between adoption and full implementation.

When students first took the new exams, significant numbers inevitably failed to achieve the required passing score. Nearly every state that implemented a graduation test first given in the eighth or ninth grade reported an initial failure rate of 30% or more.[18] Issues of equal protection and due process sparked concern among leaders of the civil rights community who fretted that the tests would disproportionately deny diplomas to black and low-income students and children in impoverished communities.[19] Black students generally passed the exams at a much lower rate than their white peers.[20] There resulted a steady stream of litigation claiming discrimination. Critics also argued that the tests lacked reliability and validity and that graduation was linked to subjective cutoff scores.[21]

Such concerns were ameliorated by the fact that every state reduced its failure rate to less than 5%-- and almost always to under 1%-- by the time the first affected cohort graduated.[22] The number of students failing the exams tended to shrink relatively quickly

as students were retested.[23] In Maryland, for instance, 75% of those who initially failed the test passed on their second try. In North Carolina, the figure was 53%.[24]

Observers disagreed about how to interpret this track record. Proponents argued that the pressure produced by MCT programs motivated students, prompted schools and districts to adjust instructional practices, focused resources on oft-overlooked students, and forced teachers to make sure they were effectively teaching basic skills. Researchers found many districts reported modifying curriculum, tutoring low-achieving students in essential skills, holding in-services for teachers on MCT, and administering pretests to students.[25] For the students who failed to meet those relatively lax standards, the most common response was remediation and repeated retesting.[26]

Critics argued that the gains were less substantive than they appeared. For one thing, more than twenty states exempted students with special needs from the requirements. More than half of the MCT states adopted achievement levels at or below the ninth grade as a passing mark for twelfth grade students. Meanwhile, critics suggested that apparent growth was largely an artifact of repeated testing. While offering repeated retests seems fair and appropriate, such a process can dilute the value of the exam—especially since many MCT programs used the same form on each administration, meaning that some gains could be attributed simply to students' increased familiarity with the test items.[27]

The Politics of High-Stakes Accountability in the States

Minimum competency testing never really went away so much as it gradually dissipated into another ineffectual educational routine. During the next decade, first the National Commission on Education's 1983 report *A Nation at Risk* and then a raft of

reformers would demand tougher graduation standards. These reformers dismissed minimum competency tests as irrelevant or counterproductive and called for more rigorous, demanding, systematic approaches to accountability. By the 1990s, efforts to promote substantive testing systems and graduation exams enjoyed widespread success and surveys suggested that the public claimed to be willing to back stiff actions—such as denying diplomas or grade promotion to students who failed to master essential skills.[28]

By 2002, more than twenty-five states had adopted mandatory graduation exams and more than twenty states offered school incentives linked to test scores.[29] As with minimum competency testing, implementation lags meant that the graduation requirements and the test-based incentives and sanctions for educators had taken effect in only a handful of states. The delays made educational sense, as they provided time to design and refine tests and testing systems and instructional content and ensured that neither students nor educators would be unfairly penalized, but they also pushed into the future the real challenges these systems would face.

The politically useful nature of the delays has been made clear as most of the handful of states that have actually started to approach initial deadlines have blinked and opted to delay the implementation of sanctions.[30] A 2000 analysis found that roughly a third of the states that have adopted high-stakes accountability systems had slowed or scaled back their original efforts.[31] In Arizona, for instance, where more than 80% of 10th graders failed the mathematic component of the state test in 1999 and 2000, the Board of Education and the legislature scrambled to push back the effective date of the graduation requirements to 2006 from the original goal of 2002.[32] In the past three years, other states, including Alaska, Wyoming, Delaware, Alabama, Maryland, and North

Carolina have also decided to scale back testing programs postpone the date at which high-stakes instruments would take effect. Other states adopted "graduation tests" but took steps to make certain that even students who did not pass could receive diplomas. For instance, Wisconsin Republican Governor Tommy Thompson pushed for a graduation exam in 1999, but union and PTA opposition to a high-stakes instrument ensured that passage would not be required for graduation.[33] In other states, such as Indiana, coercive systems were confronted with fierce legal challenges mounted by advocates for students with special needs.[34]

Because such activity does not proceed with a uniform or inexorable logic, it may be useful to consider the adoption and implementation process in a few specific states. I shall briefly discuss the challenges as they unfolded in four states where particularly visible accountability programs have been launched: Texas, Virginia, Massachusetts, and California. Such an exercise is foolhardy, as extensive analyses of the accountability regimes in each state are available elsewhere and the kind of quick survey I provide here is destined to strike informed readers as incomplete, problematic, or unsatisfying. Fortunately, the objective is not a precise rendering of developments in any one state, but to uncover patterns that may help to explain how much coercive accountability is feasible and what contextual factors shape its prospects. With that apology, whether or not it fully satisfies, I will proceed.

California

In 1977, the California legislature required all school districts to adopt proficiency standards in reading, writing, and mathematics beginning with the class of 1981. Concerned about potential backlash, the legislature encouraged local districts to only

identify a small number of competencies in each of the three basic skill areas and made clear that it was not calling for a "back to basics" approach and did not want to spur cuts in areas such as the fine arts or the humanities. Vague wording also left some ambiguity as to whether the law required that students classified as Limited English Proficient (LEP) be tested.[35]

In the initial administration of the test, 85% of white students but just 65% of black students passed the required components. Those numbers provoked concern about racial disparities, but the relatively small size of California's black population and the rapidity with which student scores increased soon quieted such concern.

In 1983, troubled by test scores and questions of quality control in the state's notoriously decentralized schools, the legislature enacted a major school reform bill that charged the Department of Education with developing curricular standards for the state's high schools. The broad, participatory effort to clarify standards was expanded to include K-8 schools and was warmly endorsed by groups including the teachers' unions and parent-teacher organizations. However, as the curricular frameworks gradually took concrete form, they conflicted with the existing tests used under the California Assessment Program (CAP). The result was an effort to augment CAP with open-ended performance tasks, requiring a scoring system that depended in part upon subjective judgment.[36]

Later in the 1980s, CAP became a victim of state politics, when Republican Governor George Deukmeijian killed the program. The death of CAP didn't have much practical effect, since its proponents had always pinned their hopes on the notion that adverse publicity produced by low test scores would compel schools to improve. That

informal approach met with little systemic success, especially since local officials argued that the content CAP tested had never been aligned with local curricula. Meanwhile, teachers were happy to see CAP expire, as it had posed a threat to the curricular frameworks that they had helped to write.

In 1991, CAP was reformulated as the California Learning Assessment System (CLAS), featuring an individualized performance-based accountability system that tested students at grades four, five, eight, and ten in a variety of subjects. The tests were to be augmented by student portfolios. The design and implementation of CLAS stirred conflict, especially among conservative parents who were concerned about the literacy and history tests. In the spring of 1994, the conflict reached a boiling point when conservative parents launched an organized effort denouncing many of the prompts used in the exam exercises as offensive. These parents cited a number of examples that they regarded as too violent, personal, or political. Representatives of groups of concerned parents wanted access to the questions. The Department of Education, concerned about preserving test confidentiality, denied them access. A court ruling that students could opt out of the tests and the state's decision to make many questions public wounded CLAS's legitimacy.[37]

CLAS suffered another blow when a 1994 *Los Angeles Times* examination of 1993 scores showed that data were skewed and the department did not follow its own guidelines in analyzing test results. For instance, while the state had developed a plan for sampling the data in which it determined that at least 25% of students at a given site would be included in school-level analyses, examiners identified more than 11,000 violations of this or other sampling rules. Leading educational groups, including the

California School Boards Association and the California Teachers Association (CTA), withdrew their support from CLAS.[38] In the end, backlash among conservative families angered by what they saw as CLAS's social agenda overwhelmed lukewarm support for a system viewed as flawed.

In 1994, the Democratic legislature passed a series of amendments intended to reform CLAS and address the complaints. Republican governor Pete Wilson vetoed the bill, but indicated he would sign on if the program provided for more emphasis on basic skills and would emphasize more traditional (i.e. multiple choice) test items. In 1995, the legislature adopted amended legislation that featured rigorous content and performance standards in core subject area and for all grade levels; an incentive program for local testing of basic skills; statewide assessment for core curriculum areas at key grade levels; and public involvement in the development of tests and the administration and reporting process. Wilson signed the bill over strong conservative opposition.[39]

The legislature superseded the 1995 program just two years after its launch, approving in 1997 the Standardized Testing and Reporting (STAR) program. STAR included a new basic skills test for students in grades two through eleven, the adoption of state standards, and the development of an exam based on those standards. While appropriate tests were devised and refined, however, officials had five weeks to choose an interim test.[40] They opted to use the Stanford Achievement Test, ninth edition (SAT-9).[41] The hurried implementation led to a number of public difficulties. The most remarked upon was the fact that in 1998 nearly 700 of the state's 8,500 schools got inaccurate test results and more than 750,000 students were omitted from the statewide

analysis.[42] Testing opponents seized on such incidents as evidence that high-stakes testing was capricious and unreliable.

As Mike Kirst has noted, "The Stanford 9 rapidly became the tail that wagged the accountability dog. As the accountability system continued to unfold under the next governor, [Democrat] Gray Davis, monetary incentives for teachers and schools were attached to test-score gains on the Stanford 9. Schools got the message and began preparing students for the material tested by the Stanford 9—even though they were supposed to be teaching the state's curriculum in order to prepare for the California-developed assessments that were to come later."[43]

In spring 1999, the legislature adopted another ambitious accountability measure, the Public Schools Accountability Act (PSAA), to complement existing efforts. The Act included three major components: the Academic Performance Index that would measure and rank school performance, the Immediate Intervention/Underperforming Schools program that would target resources to poorly performing schools, and the Governor's Performance Award program to give cash bonuses and other incentives to schools and teachers whose students fared particularly well.[44] Taking advantage of a large state budget surplus, Governor Davis and the legislature directed the state to move promptly in providing new funding to low-performing schools and awarding bonuses to high-performing teachers and schools. Perhaps unsurprisingly, given that the state was using the carrot rather than the stick, it was the first time in California's long history of accountability that provisions were rapidly implemented. Even that effort created backlash from some CTA officials and classroom teachers, who saw the incentives as the leading edge of an effort to sew division among the state's educators.

The high school exit exam that would take effect with the class of 2004. The High School Exit Examination (HSEE) was first administered in spring 2001 to ninth graders, who were to be permitted to retake the test each subsequent time it is offered until successfully completing each section. The legislation required that English Language Learners and students in special education pass the HSEE, although districts may defer testing some ESL students for up to 24 months and some students with special needs may receive accommodations.

In December 2000, responding to concerns over student preparedness for the exit exam and the length of the proposed 200-question exam, the State Board of Education voted to shorten the test by eliminating some of the more difficult algebra questions and to eliminate some multiple choice questions from both the reading and mathematics portions. The changes reduced the number of algebra questions from 26 to 12 and the number of language arts questions from 100 to 82. The Board's decision was supported by Governor Davis and the Superintendent of Public Instruction.[45]

Early in 2001, confronted with the prospect that large numbers of students would fail and with continuing CTA hostility to test-driven sanctions, legislators took up proposals to delay implementation of the graduation requirement. In January 2001, CTA President Wayne Johnson reiterated the union's concerns about high-stakes exams while discussing the Bush administration's education proposals, saying, "[Standardized] tests should not be the sole criteria for determining what public school students and teachers are really accomplishing."[46] In February, the state Senate voted to push back the graduation requirement by a year, from 2004 to 2005 and to make the inaugural administration a practice test for ninth graders. The measure was reversed in the

assembly, however, where legislators wanted to see the results of the initial trial in spring 2001 before taking action. Meanwhile, some legislators made it clear that they thought even 2005 would be too soon to implement the graduation requirement and that more substantial delay would be necessary.[47]

The new test was first administered in March 2001, amid confusion about whether this was a "practice" run. About 400,000 students, or 81% of the state's 9th graders took the test and more than 55% of those failed. Superintendent of public instruction Delaine Eastin termed the results "sobering" and acknowledged "the data show that we have a great deal of work to do."[48] Among Hispanic and black students, the failure rate was over 75%. In a state whose 2000 population was 32% Hispanic and 7% black, such failure rates among the minority population were clearly untenable and triggered the stirrings of organized unrest. During 2001, the anti-CLAS Coalition for Education Justice sponsored several rallies and teach-ins to protest the test, drawing 300 people to a Los Angeles rally where educational officials were urged to protect students from "racist and class-biased high-stakes testing."[49] Opposition also emerged in some wealthy communities where teachers and parents viewed the tests as intrusive and interfering with the development of "critical thinking, class interaction…and the development of ideas." For instance, a school board member in California's Marin County urged parents to boycott the state tests, prompting parents of about 600 of the district's 2,700 students to provide waivers excusing their children from the tests.[50]

Meanwhile, districts are working to boost passing rates. History suggests that they will succeed. Districts are implementing professional development, incentives to lure strong teachers into low performing schools and improve assessment scores, and

summer school programs for students who fail the exam. In California, where an influential teachers' union, powerful civil rights organizations, conservative opposition to the kind of holistic assessments that the CTA prefers, and a Democratic legislature create a hostile milieu, coercive accountability has historically failed to gain much traction. In 2002, hostility between the Governor and the CTA and a rapidly growing state deficit have put the test-driven bonus payments at risk. Whether the current effort can take root in such an environment is not yet clear.

Massachusetts

The Massachusetts Educational Assessment Program (MEAP) was established by the Massachusetts School Improvement Law of 1985. Signed into law by Democratic Governor Michael Dukakis, and enacted by an overwhelmingly Democratic legislature, the law mandated that the MEAP be administered biennially but attached no consequences to test results. First administered in 1986, the MEAP was retired in 1996 in favor of the new Massachusetts Comprehensive Assessment System (MCAS).

Comprised primarily of multiple choice questions, the MEAP was designed to provide information that would help to improve curriculum and instruction and permit comparison at the school, district, and state levels. The test did not provide individual student results and the state did not establish a passing score. The system was intended as a diagnostic and pedagogical device—thought some advocates did hope that mediocre scores might provoke public unrest.

In 1993, Massachusetts enacted the ambitious Massachusetts Education Reform Act (MERA) to replace the old MEAP, substituting a potentially coercive vision of standards-based reform for the previous diagnostic regime. Championed by liberal

gubernatorial hopeful Mark Roosevelt, the MERA was a response to massive inter-district disparities in funding and performance. Promising public accountability and clear benchmarks for student achievement, proponents argued that students would no longer be passed through school systems without acquiring basic academic skills. While such an approach held natural appeal, the reformers also extended an olive branch to accountability opponents by including legislative language that called for the need measure student learning in multiple ways.

The 1993 legislation mandated changes in curriculum and instruction, teacher preparation, student assessment, governance and decision-making, and education finance. The pivotal link in the system was the MCAS assessment. Students in tenth grade were tested in four areas: English, Math, History, and Science. Starting in 2003, students would have to pass the tenth-grade mathematics and English tests to graduate.[51] Most special education students were expected to take the MCAS with some accommodations; a few would take an alternate MCAS.

In 1996, Republican Governor William Weld championed a successful effort to overhaul the Massachusetts Board of Education, transforming it from an unwieldy 17-member body into a less insulated nine-member body. Weld appointed as board president the controversial and hard-charging John Silber, a strong advocate of high-stakes accountability who had been criticized as autocratic during his tenure as president of Boston University.

In 1997, a poll of Massachusetts residents conducted found that 61% supported passage of a 10th grade competency test as a condition of high school graduation. While about half of those expressing an opinion thought that no more than 10% of the students

in their own communities would fail, more than sixty percent said that they would still support it even if 25% of their hometown students failed the exam.[52] The question was how many students would actually fail, and what would happen to that support when the hypothetical students were one's own children, or those of friends and neighbors.

The MCAS was first administered in 1998. The test was billed by the Board of Education as more rigorous than the typical state assessment. In fact, when the Board set the passing scores required for graduation in January 2000, board members opted for a standard (a threshold scaled score of 220) that roughly half of the state's tenth graders had failed to meet just two months before.[53] Even that standard was derided by critics as representing "a low 'D' level passing grade."[54] Nonetheless, the Board found it easier to set a high threshold than elected officials would have, especially in the case of Massachusetts where the leaders of the heavily Democratic House and Senate enjoyed close relationships with the Massachusetts Teachers Association (MTA). As one MTA official noted in a private communication, "We have relatively little influence over the Board of Education, which is appointed, not elected. A lot of education policy is made by the board, which establishes regulations pursuant to legislation."

Failure rates on the initial test varied tremendously by race. In 2000, 61% of black students and 66% of Hispanic students failed the tenth grade English language arts test, while just 28% of white students failed. In math, the results were even more stark, with 76% of black students and 80% of Hispanic students failing, compared with 38% of white students.[55] The math failure rate alone meant that more than 15,000 white students were in danger of being denied diplomas. The number of failing black and Hispanic students was far smaller only because the fewer than 8,000 minority students even took

the test. While denying diplomas to more than 50% of black and Hispanic 12th graders was clearly unpalatable, the small size of the state's minority population made the results less damning than they would have been in California.

These results generated heated opposition to the test among teachers, civil rights organizations, and liberal activists ideologically opposed to the test regime. In particular, the state's staunch liberal communities—especially Cambridge and the elite Boston suburbs—provided fertile ground for an array of anti-test organizations. Comparing themselves to the "Freedom riders" who resisted Southern segregation laws, a "Committee of 100 Massachusetts Parents"—composed primarily of Boston-area parents—tried to organize a boycott of the exams. By 2001, other groups such as the Students' Coalition for Alternatives to the MCAS (SCAM) and the Coalition for Authentic Reform in Education (CARE) were also holding rallies around the state, promoting boycotts, and lobbying officials to revise or dismantle the MCAS.[56]

During 2000-2001, the Massachusetts Teachers Association launched a $600,000 advertising campaign that attacked the "one-size-fits-all, high-stakes, do-or-die MCAS test." The administration of the state's Republican governor Paul Cellucci responded by directing the state to launch an aggressive, $500,000 television and radio ad campaign on behalf of the exam.[57] The Cellucci administration also proposed a number of modifications for the MCAS that included expanding the allowable testing accommodations for students with special needs, narrowing the world history section to focus on American history, and permitting students who had not passed by the end of 12th grade to enroll in alternative programs at community colleges.[58] A spring 2001 poll of 300 teachers conducted by the Boston Teachers Union found that about 85% of the city's

public school teachers opposed using the MCAS exam as a graduation requirement. Just seven percent of teachers backed the requirement.[59] In May 2001, the MTA also released a poll it had commissioned from Kiley & Co. that claimed 54% of Massachusetts residents now opposed the MCAS graduation requirement—even if students were permitted to take a "scaled-down" retest.[60]

In 2001, more than four dozen bills seeking to modify the MCAS were filed in the state legislature. The proposals sought to do everything from creating exemptions for certain kinds of students to repealing the tests outright. More than a 150 people, nearly all hostile to MCAS, spoke during a day-long hearing on the topic. Despite the rumbles of concern from the anti-test parent groups, the civil rights community, and the state's teachers, the proposals were largely shrugged off by the heavily Democratic legislature.[61]

This resolution was somewhat surprising given that union opposition and the possibility that large numbers of students would not qualify for diplomas generated pressure to make accommodations. What explains this outcome? In part, demands on the legislature were lessened because the governor, the Department of Education, and Board of Education made it a point to bend in response to several concerns.

In late 2000, Governor Cellucci announced a plan that would permit students with disabilities to receive a Certificate of Completion without having to pass the MCAS. At about the same time, the Board of Education opted to delay making the science and history MCAS tests part of the graduation requirement. In January 2001, the board voted to give students five chances to pass the tenth-grade MCAS test and to omit some of the hardest questions on the retests.[62] Some board members also suggested the board ought to contemplate an alternative test. Education Commissioner David P Driscoll explained,

"Let's say a kid can demonstrate, with a hands-on approach, his knowledge of geometric concepts. If he can demonstrate in another way, we should at least explore that."[63]

In spring 2001, departing Board of Education member Ed Delattre blasted MCAS and called for the graduation requirement to be delayed until 2009. Delattre, dean of the Boston University school of education, took pains to explain he did not oppose accountability, only that he feared, "We are nowhere near being able to guarantee that all students have been the offered the opportunity to meet the academic learning standards MCAS should be testing."[64]

In January 2002, the Board of Education unanimously adopted an appeals process that would permit students to graduate without passing the MCAS exams if they could otherwise prove they possessed the requisite knowledge. Education Commissioner Driscoll noted, "This is simply an issue of fairness. Some students, for whatever reason, cannot demonstrate their real level of performance on MCAS."[65] An education department spokesman predicted that about 2-5% of seniors statewide would be eligible for the appeals process.[66] Later in the spring, the Board and Department of Education officials backed away from a plan to include MCAS scores on high school transcripts, reducing the stakes riding on test performance for high-achieving students.

In April 2002, after two rounds of test results, the state Department of Education reported that about 15,300 juniors still had to clear the MCAS hurdle. But state officials focused on the fact that 76% of juniors—or about 48,400 students—had successfully passed a graduation test regarded as one of the nation's toughest. That figure was up from 68% prior to the first retest, though two-thirds of students failed the math retest and slightly over half failed the English retest.[67] Students have five chances to pass the

English and math portions before they graduate. Meanwhile, test opponents focused on the fact that—in spring 2002—a quarter of the class of 2003 was at risk of being denied a diploma at graduation.[68] Critics worried that many of the students who still needed to pass a test would get over the bar only by focusing narrowly on test preparation at the expense of substantive learning.[69]

During 2001-02, the state also backed efforts, such a model program sponsored by Holyoke Community College, to support students who failed the graduation exam. The Holyoke program paired the college with local high schools to offer students intense preparation, career counseling, and a chance to take college courses while preparing to re-take the exam. Programs like this, which permitted students to receive college credit and move on with their lives even if they failed the MCAS, had the politically desirable effect of softening the blow of test failure—and did so especially for those low-achieving students who demonstrated particular concern for schooling.[70]

Massachusetts, a state with an active political tradition and a legislature sympathetic to inequities, poses a challenging test for proponents of high-stakes accountability. The MCAS system has received national praise, student scores have shown dramatic improvement, while the state Board has embraced a number of refinements that have slightly lowered the bar, permitted some students to side-step the test, and have softened the consequences of failure. Whether these refinements will stabilize and protect the system if thousands of Boston-area students are denied diplomas, or whether they will later appear to have been the initial steps as the state edged away from coercive accountability, is an open question.

Texas

In 1979, the Texas legislature enacted the Equal Educational Opportunity Act, which established the state's first testing program. It required the Texas Education Agency (TEA) to create a series of criterion-referenced assessments to assess basic competencies in reading, writing, and mathematics. The TEA responded by designing the Texas Assessment of Basic Skills (TABS).

In 1983, the legislature mandated that students who failed TABS would have to retake it each year. There were no consequences for failure, but the legislature hoped that requiring students to retake the exam would highlight problems and pressure schools to provide remedial support for students in need. As part of the effort to drum up pressure on schools and teachers, test results for each district and school were made publicly available for the first time.[71]

In 1984, the Texas Education Code was amended so that it referred to "minimum" basic skills rather than "basic skills competencies." Supported by State Board of Education, TEA officials interpreted the change as a requirement that they make assessments more stringent and begin to sanction students for inadequate performance. Legislation changed the revised eleventh grade assessment to a graduation test starting with the class of 1987.

Exit exams were first administered in October 1985 to 190,000 eleventh-graders. Eighty-eight percent of students passed the math portion, 91% passed the English/language arts portion, and 85% passed both parts. Proponents of rigorous accountability assailed the tests as being too easy, arguing that the impressive passing rates were a product of overly simple assessments and a low passing threshold. Students

who failed either or both parts were allowed to retake the tests in May 1986. The majority of students who had originally failed passed the spring retest.[72]

In the late 1980s, the State Board of Education directed the TEA to boost the academic rigor of the exams and increase the curricular validity of the assessments by linking tested content more closely to the state's core curriculum. The new assessment program, the Texas Assessment of Academic Skills (TAAS), was implemented in 1990. It sought to shift the testing focus from minimal skills to higher-order thinking and problem-solving skills.

In 1990, TAAS was administered to students in grades three, five, seven, nine, and eleven, with the eleventh grade test serving as the new exit exam. On the fall 1990 TAAS tests, students fared far worse than they had on the previous TEAMS tests. If the Board had retained the 70% passing score from the TEAMS test, passing rates would have declined from the 80-90% to the 40-60% range. Moreover, passing rates for black and Hispanic students on the math portion would fall to about 30%.[73] However, the State Board had decided to set the passing score at 60% rather than 70%, opting to phase into a 70% threshold over a two-year period.

Over 165,000 students had taken the test, and a staggering 38,000 failed to meet even the new sixty percent passing standard. After two additional rounds of retesting, Education Commissioner Lionel Meno reported that the number of students unable to pass had shrunk to a far more manageable 7,996. Those students did not receive diplomas in May 1991, although nearly a quarter subsequently received them after finally passing during a summer administration. In 1992, the State Board of Education considered not moving forward with the plan to raise the passing score to 70%, but

finally decided to go ahead as originally planned. Of the first cohort of eleventh-graders subject to the seventy percent cut-off, 51,000 out of 187,000 (or more than 25%) initially failed. However, retesting and targeted support shrunk the number of students denied diplomas to about 5%.

In 1993, with public pressure for enhanced school accountability—especially from influential business groups like the Texas Business and Education Coalitions—the legislature enacted an accountability system that linked school- and district-level incentives to student TAAS performance.[74] The legislature mandated that student performance data be disaggregated into African-American, Hispanic, White, and Economically Disadvantaged and required schools to perform effectively in each subgroup, and to meet certain other criteria relating to drop-out rates and so on.. "Exemplary" schools had to have at least ninety percent of students in each subgroup pass each subject area. To be rated "acceptable," schools were to have at least 40% passing rates in each group, a figure that was gradually ratcheted up in ensuing years. Significantly, the 1993 legislation put into place a series of sanctions for schools and districts where student performance failed to meet guidelines. Schools deemed "unacceptable" had three years in which to improve performance, after which they could be subject to state takeover or forcible closure of a school.[75]

In spring 1994, testing was expanded to additional grades and the exit exam moved to grade ten from grade eleven. Moving the graduation test to grade ten allowed more time for schools to remediate and retest students, but also required scaling back test content. During 1990 to 1994 period, high school tests in Algebra I, Biology, English II, and U.S. History were added. Students had the option of passing the algebra, English,

and either the biology and U.S. history tests as an alternative to the TAAS graduation test requirements, weakening the assurance that students had mastered a particular body of knowledge and skills but gratifying parents and teachers who feared that substantive instruction in these advanced courses was threatened by a focus on the graduation tests.

Schools initially fared poorly on TAAS. Just 53% of students achieved passing scores. The figure was just 31% for black students and 39% for Hispanic students. In Texas, the combination of limited student-level sanctions, gradually stiffening school-level accountability, and weak union opposition helped the accountability system overcome the opposition produced by these initial results. As one journalist noted in 1999, many observers suggest that Texas's reform efforts have benefited from, "Texas' lack of strong teachers unions, which…lets reformers make change quickly, but ensures that such change can never be replicated nationally without union-busting coast to coast."[76]

About 40,000 students in total were denied diplomas between 1994 and 2001, but the numbers fell steadily each year.[77] By 2000, the overall passing rate had climbed to 80%, including 67% for black and 70% for Hispanic students. In 2001, just 3,723 seniors were denied diplomas based on test scores. Ninety-five percent of white students passed the TAAS graduation test, while 84% of Hispanic and 82% of black students did.[78] Moreover, schools were excluding a declining share of their students from testing, even as overall performance was improving and the racial achievement gap shrinking.[79]

Critics claimed that the improvements did not reflect increased learning, charging that the steadily improving test scores were largely driven by schools holding students back, schools trying to exclude special education students from testing, increased drop-

out rates, and a gradual easing of the number of questions students had to correctly answer to pass.[80] Some analysts also suggested that test scoring had been eased over time, with critic Walt Haney arguing students had needed to get 70% of questions correct in order to pass throughout the 1990s but that the figure had slowly dropped to about 50% in 2000.[81] State officials countered that such adjustments were appropriate given the increased rigor of the tests, but they also show how inescapably arbitrary are some essential decisions.

Other analyses suggested that TAAS did include substantially fewer tough questions than high-stakes tests in other states such as New York, Kentucky, and Massachusetts.[82] State officials acknowledged the critique, but pointed out that the test had been adopted a decade before and planned to launch a new, more rigorous test to address the problem.

Whatever the merits of these various critiques, without a strong teacher union or civil rights community to trumpet their case, the critics enjoyed little success in their efforts to soften TAAS. In 1999, the legislature directed the TEA to begin developing a new testing program intended to align objectives for all grade levels of the state tests. To be called the Texas Assessment of Knowledge and Skills (TAKS), the new regime is to replace TASS beginning in 2003. TAKS is an expanded set of state tests that will be linked to the tougher state standards adopted in 1997.

Beginning with the class of 2005, students will be required to pass eleventh grade exit tests in mathematics, English, science, and social studies. Those students who do not initially pass the exams will have up to eight additional opportunities to do so. The transition to the new system promises the same kind of difficulties that occurred during

the TEAMS/TAAS changeover. In December 2001, Texas Commissioner of Education sent a letter to school districts that indicated—based on current student scores—just 44% of students would pass the new exit exam.[83] Students in lower grades will also begin to face high-stakes assessments. Starting in 2002-03, third grade students who do not pass the TAAS reading assessment after three tries will not be automatically promoted. By 2007-08, similar policies will be in place for fifth-graders and eighth-graders in the subjects of both reading and math.[84]

The Texas experience has been one of unusual stability. In a state with weak and fragmented teacher unions, an organized and influential business community, a relatively conservative legislature in which civil rights organizations and the educational establishment historically enjoyed limited sway, the broad commitment to accountability held up even in the face of occasional opposition. However, that stability also benefited from design decisions in which retesting provisions ensured that the number of students initially denied diplomas was manageably small and a scoring system was designed that was lax enough that schools and districts fared acceptably.

Virginia

In 1976, Virginia became one of the first states to embrace the budding minimum competency movement, despite the opposition of established civil rights and educational organizations such as the Virginia Education Association (VEA) and the National Association for the Advancement of Colored People (NAACP). The opponents never mounted a very effective fight, in large part because Virginia is not a collective bargaining state. As in Texas, this sharply curtailed the impact of teachers' union—typically the most effective opponent of coercive accountability.

Initially, the standards that accompanied minimum competency testing were billed as an innocuous effort to "clarify" the state curriculum. The launch of the test was delayed because state officials were unable to agree on what the passing score should be, before the passing score was finally set at 70% in 1978.[85] When the test was first administered in 1978, 17.8% of the 71,000 10th grade test-takers failed either the reading or mathematics component. Forty-two percent of black students failed at least one component, prompting the executive director of the Virginia NAACP to blast the test.[86]

Criticism subsided, however, as the passing rate of black students rapidly rose, along with all other students. In 1981, just 0.5% of black students failed the exam, while the state denied diplomas to just 87 of the 62,236 seniors who took the test (a denial rate of 0.14%).[87] Soon, concern emerged that the standard was too low to be effective.

In 1986, Democratic Governor Gerald Baliles endorsed a new "Literacy Passport Test" (LPT) to ensure that all sixth-graders were performing at an acceptable level in reading, writing, and arithmetic. No student would be promoted to ninth grade without passing the test. Special needs students were exempted from the program. In 1990, the first year of LPT testing, 71% of white test-takers and 46% of black test-takers passed the LPT. In 1991, the passage rate among black students improved to 53%, but still lagged the white rate by 26 percentage points. The results provoked outrage among black leaders.[88]

By 1992, of the initial cohort of test-takers, 5,000 had been promoted to eighth-grade without passing the LPT, despite the policy that had initially prohibited such promotions. Officials also removed or lowered the testing bar for some students (like ESL students) who would have difficulty passing the exam. In the end, the number of

BEST COPY AVAILABLE

students denied diplomas due to the LPT was negligible. In 1996, just 83 students were denied diplomas due to their LPT scores. In 1997, just 148 students were denied diplomas. The 99%+ pass rate was reminiscent of the experience fifteen years earlier with Minimum Competency Testing. The next year, the state decided to phase out the LPT, deeming it unnecessary in light of the newly adopted Standards of Learning.

During the 1980s, Republicans made substantial gains in the Virginia legislature. By the time Republican George Allen was elected governor in 1993, as the first Republican governor in twelve years, Republicans held more than 40% of the seats in the legislature.[89] Whereas Democratic legislators enjoyed substantial support among the groups most likely to critique or oppose high-stakes testing—for instance, the 50,000 member VEA and the NAACP—Republican legislators were more willing to support measures these groups opposed.

At Allen's behest, in May 1994, the Board of Education initiated the development of statewide standards in math, science, English, and history. While crafting the standards in math and science was relatively consensual, there was some disagreement over the relative emphasis that the language arts standards ought to devote to phonics, and significant conflict over the proposed social studies standards.[90] These fights were the same ones that plagued California's accountability efforts at about the same time. Critics, including the Virginia Education Association and the Virginia Association of School Superintendents, accused the administration of having rewritten the standards to reflect a more conservative perspective and of desiring social studies and language arts standards that promoted the "regurgitation of isolated facts" and "lower-level thinking

skills."[91] Debate over the merit of these charges would continue, leading to various efforts to adjust the standards and address the critics.

During 1994 and 1995, SOL proponents engaged in a delicate dance. They touted the virtues of meaningful standards without inflaming opponents who opposed holding educators accountable for student performance. Board members publicly promised that the board did not envision "hold[ing] any teacher accountable for any test score" and that the board had no intention of using test results punitively against school divisions or individuals.[92]

The SOL tests, administered for the first time in 1998, were criterion-referenced tests designed to measure whether students master the content specified in the state curriculum. The tests consisted entirely of multiple-choice questions in all subject areas except English. Once state graduation standards took effect in 2004, students would have to pass six of the twelve "End of Course" high school exams to earn their diploma.

In October 1998, with the first administration of the SOLs looming, the Board of Education had to set passing scores for the tests. The bar-setting exercise provoked fierce conflict in pro-SOL ranks, as some board members were criticized or having "gone soft" when they supported cut-off scores that hard-liners deemed too low. Those favoring moderate thresholds attacked the hard-liners, in turn, for demanding unreasonable standards that would demoralize students and educators.

When huge numbers of students failed the SOL tests in 1998 and 1999, pressure grew to create a safety valve to accommodate them. In 2000, Board president Kirk Schroder proposed a "basic diploma" for students who passed the English and math tests and demonstrated that they possessed job skills but failed to pass the half-dozen high

school SOLs required for a standard diploma. Critics feared that the proposal would create a two-tiered educational system.[93] In the end, the proposal was dropped.

While the Board dropped the basic diploma proposal, it did create a "modified standard diploma" allowing the 14% of Virginia students enrolled in special education to bypass the SOL tests.[94] Accommodations were also made for students with limited English proficiency (LEP), permitting them to opt out of one year of SOL testing, to take the tests using a bilingual dictionary, and specifying that the scores of such students would not count towards the school composite for two years.

In 2000, the Board also responded to protestations from some of the state's high-achieving districts, who complained that the SOLs were interfering with curricula and instruction in elite classrooms. The Board permitted high school students to substitute such board-approved tests as the Advance Placement (AP) or SAT II for the appropriate SOL test. This approach offered succor to those high-performing teachers and students most likely to feel straitjacketed by the SOLs.

Recognizing that holding educators responsible for student test performance would provoke opposition, reformers at first showed little inclination to link school evaluation to student performance. Eventually, in 1997 the Board adopted a performance-based accreditation system, although the new requirements would not take effect until 2006-07 and board members explicitly ducked the question of what it would mean for a school to lose its accreditation. The only specified consequence for failure to meet accreditation standards was that schools would have to adopt "a three-year School Improvement Plan." The Board did not specify what would occur if the plan did not produce the desired results.

SOL critics argued it was a mistake to rely too heavily on SOL performance in evaluating schools or students.[95] In the 2001 legislative session, several "multiple criteria" proposals emerged that sought to base graduation on more than SOL test performance. Board of Education president Kirk Schroder criticized the proposals, terming them "simply a back door to help students who otherwise should have failed the tests."[96] None of the proposals passed.

In the first round of SOL tests in 1998, Virginia's schools showed abysmally. Just 39 of Virginia's 1,800 public schools had satisfied the 70% passing rate required for school accreditation. More than 97% of the state's schools were out of compliance with the new standards. Critics, such as a former president of the Virginia Educational Research Association, took this as evidence that "the SOL test results misrepresent the condition of public schools in Virginia."[97] SOL proponents countered that the results illustrated the mediocrity prevalent in Virginia's schools.

In 1999, test results improved substantially—though they remained abysmally low—as 6.5% of schools had at least 70% of students pass. Officials and educators wrestled with whether they ought to celebrate the tripling in the percentage of satisfactory schools or bemoan that more than 90% of the state's schools were still failing. Black student performance improved on 26 of the 27 SOL exams and the black-white gap closed on 16 of them. Nonetheless, while 41% of white students failed one or more tests, three-quarters of black students did. In 2001, results were again up significantly. In 2001, 40% of schools now performed well enough to meet accreditation standards, a fourfold increase from 1999. By 2001, more than 80% of high school students were passing the English SOLs and more than 70% were passing the algebra and geometry

SOLs. Those numbers were up sharply from 2000 and, in the case of math, marked enormous 21- to 43-point increases from the initial 1998 results. In 2001, black students again made dramatic strides, but their scores would still result in more than half of them being denied diplomas were the graduation exams in place. Similarly, despite the impressive gains, 30% of schools still were not meeting even provisional benchmarks for accreditation—though it still was not clear what that meant in practice.

After the SOLs were launched in 1998, implementation sparked resistance, even though it would be 2004 before student results had consequences and 2007 before school results did. In March 1999, SOL critics launched Parents Across Virginia United to Reform SOLs. By fall 1999, its membership numbered 2,200. Critics argued that SOL scores did not reflect real gains or were due to an unhealthy focus on testing and test preparation, while claiming that they were not opposed to accountability in principle—only to the SOLs as currently designed.[98]

Even as test scores jumped between 1998 and 2001, doubts about the program remained and opposition grew in some quarters. An August 2000 Washington Post survey of registered Virginia voters found that 51% said that the SOL testing program "is not working" and 34% said it "is working." Asked what should be done about the tests, 43% said they should be substantially changed and 21% said they should be "ended entirely." Just 24% of respondents said they should remain "as is."[99] An October 2000 *Richmond Times-Dispatch* poll of registered voters found similar discontent with the SOLs, with 66% saying that SOLs were not the best way to measure student performance and 68% that they were not the best way to measure school performance.[100]

Despite the unrest, Virginia's Republican legislature remained firmly committed to the SOL program as of early 2002. Mark Warner, the first Democrat to win a gubernatorial election in Virginia since the high-stakes SOL system was introduced, committed himself to the program during his 2001 campaign. Meanwhile, there was some evidence that parents, educators, students were becoming acclimated to the system, grumbling about it but also accepting it as fact of life. It is not yet clear whether this early stability will hold up as the program's sanctions take effect.

How Context Matters

The politics of coercive accountability is a clash between aggrieved groups with concentrated interests and a broader public that stands to reap diffuse benefits. Such fights generally have a predictable calculus, whether in the case of agricultural subsidies or military base closures, with the concentrated interests emerging triumphant. However, the outcome of any specific conflict depends on the backdrop against which it plays out. When the aggrieved interests are larger, more influential, or more organized, they will be more successful at fending off efforts to promote coercive accountability. On the other hand, when such groups are weaker or have less purchase on the decision-makers or when coherent interests emerge to champion accountability programs, coercive programs are more likely to survive largely intact.

Consideration of developments in the various states suggests a number of hypotheses as to how seven key contextual factors may shape the outcome of this conflict. Obviously, the roles of the seven have not been rigorously examined here. Such analysis will require further efforts that more systematically consider how they influence the fate of transformative accountability.

State racial composition. Because poor performers are concentrated in low-income, minority communities, it is the leadership of ethnic groups or the civil rights leadership that often lead the attack on the equity of coercive arrangements. In states like Massachusetts, where the minority population is relatively small, proponents of transformative accountability will face less opposition than in a state like Virginia where there exists a large black population with a shared history of deprivation. The larger the base of support that these leaders represent, and the more votes they appear to represent, the more pressure officials will face to soften sanctions. Given the historic nature of black deprivation in American education, the symbolic clout of black opposition can prove especially potent when wielded by active leaders or when joined with a strategy of aggressive legal contestation.

Teacher unions/associations. Due to their desire for classroom autonomy and job security, leaders of teacher organizations are generally hostile to coercive arrangements. The power of their opposition depends on the political muscle of their organization and the institutional levers they control. In collective bargaining states, like California or Massachusetts, unions wield significantly more influence than in states such as Texas or Virginia. The stronger the union in a given state, the more difficult it is for proponents to advance and hold the line on coercive arrangements.

Ideological or neighborhood communities. Opposition to high-stakes accountability is often rooted in concerns that the public schools are being pushed to teach skills or behaviors that families find morally troublesome. In California, fundamentalists attacked the hidden agenda of the testing system; in Massachusetts, liberal activists attacked the MCAS for stifling creativity and perpetuating racial and

socio-economic inequities. Opposition has also often emerged among wealthier communities fearing the pressures of testing are undermining advanced courses or that test-driven school evaluations are unfairly hurting the reputation of their local schools and the value of their homes. The impact of either of these groups will depend upon their presence and upon the degree to which they are organized and vocal.

Business community involvement. The business community is one coherent constituency likely to mobilize on behalf of coercive accountability. When the business community is sufficiently concerned about school quality, its organizations and partnerships may systematically work to advance coercive accountability. Backed by the resources and support of leading business interests, and aided by the perception that they are focused on advancing the public interest, such groups can encourage public officials to hold the line on the more troublesome elements of high-stakes accountability systems. In states with a business community that has historically played an active role in education reform, such as Texas, the activity can counter opposition from the irate.

Partisan makeup. Generally, Democratic officials are more reliant on the support of public sector employees and minority voters than are Republican officials. Consequently, the balance of power held by active interests rests in part on the partisan backdrop. In particular, states with a strong Republican presence will find it easier to resist the opposition raised by teacher unions and civil rights organizations. However, when Republicans represent suburban communities that enjoy highly regarded school systems they may face constituent pressures to soften systems that are thought to be stifling classrooms or unfairly tarring effective schools. If substantial blocs of

Republicans represent such districts, the legislature will not necessarily prove receptive to coercive reform.

Existing system. The greatest obstacle for coercive accountability systems is the perception that they are unfairly punishing some students or educators. Such perceptions are rooted in experience. If over the past decade a state has routinely recognized schools based on student performance, has denied diplomas to 12th grade students who failed the exit exam, or has linked administrative pay to demonstrated student outcomes, these practices come to be seen as an established fact of life and rarely occasion much discussion. It is in the introduction, implementation, or toughening of such systems that backlash emerges. Consequently, the more experience a state has with some version of high-stakes accountability, the easier it will generally be to erect a system of coercive accountability. The efforts in Virginia and Texas, for instance, benefited from the ability of legislators and officials to erect each successive testing regime on the foundations of the previous effort. While such cyclical reform produces its own problems, existing public receptiveness to test-based sanctions makes such efforts easier.

Boards of Education. Education is unusual in that many states feature strong, quasi-independent Boards of Education that govern, to various degrees, important elements of K-12 schooling. Active boards can serve to insulate governor and legislature from direct discontent or fallout. Because the key decisions are often being made by "independent" appointees, it can both foster a greater ability to withstand political discontent and create more room for compromise—because the arrangement permits board members to take stands that the governor would be skewered for taking. Board involvement can also cut the other way. Boards can readily enhance the viability of a

coercive system by quietly tweaking "technical" aspects like cut-off scores, question content, and retest opportunities in a way that substantially raises passing rates. The direction of board involvement will generally turn on whether members see themselves as there to ensure the political success of the accountability regime or to ensure that it retains its coercive promise.

Conclusions

High-stakes accountability is effective to the degree that it is coercive. When jobs or diplomas are at stake, educators can no longer so readily close their classroom doors and wait out reforms. They are driven by self-interest to alter their practice in the intended fashion. From the inception of high-stakes testing, proponents tend to laud the requisite tests as clear, scientifically defensible, manageable, and concise. Critics typically attack them as unreliable, simplistic, too brief for their intended purpose, overly focused on trivia, or lacking the necessarily curricular and pedagogical support.

In seeking to answer these concerns, all of them legitimate to some degree, proponents try to tweak systems without refining them into irrelevance.[101] For instance, adjusting required scores or giving students multiple chances to pass the test can be a useful and appropriate exercise, or can risk undermining the very purpose of the transformative accountability. Adding essay questions can usefully broaden assessment; it also renders scoring more subjective and can sometimes make the cost of testing prohibitive. Giving students five or eight chances to pass a test can ensure that no one is denied a diploma due to ill-fortune and gives students the incentive and opportunity to improve the performance; it can also undermine the system by permitting some

substantial number of students to slide through based on the one test they took where they caught all the breaks.

The effect of such "refinements" depends largely on the context in which they take place. The crucial component is the willingness of a majority of voters and officials to tolerate state sanctions on students or educators. It is far easier to build a stable, rigorous accountability system when the public will shrug off 5,000 students denied diplomas or fifty schools reconstituted than when it will accept only a fraction of that number. This is essentially the same kind of sensitivity to the public's willingness to accept military casualties that constrains national security officials as they consider military interventions.

At any given level of public resolve (or callousness, depending on one's perspective), however, there are also institutional and behavioral factors that may help or hinder efforts to erect substantive accountability systems. Particular attention is called to seven factors: the strength of teacher unions, the minority community and civil rights organizations, ideological communities and resistant suburban enclaves, and the business community; the partisan makeup of the state and the legislature; the character of the pre-existing accountability system (if any); and the strength and independence of the state Board of Education. Discussion of the effects of these factors is, at this point, preliminary and tentative. There is extensive room for scholarship that systematically considers these factors, the role they play, how they interact, and the implications for public policy.

The decision to embrace high-stakes accountability represents a choice to trade a system in which each child's education depends heavily upon the skills and outlook of

her teacher for one characterized by standardized norms and measures of performance. It means swapping the strengths and frailties of an education system reliant on goodwill and intrinsic motivation for one anchored in the firmer ground of self-interest. While alluring in the abstract, this trade-off clashes with the values that permeate traditional public schooling and inflicts heavy costs on particular constituencies. The result is a tendency to recoil from the reality of standards, resulting in a series of well-intentioned compromises that leave the façade of accountability intact but strip its motive power.

Proponents have difficulty standing firm on the details of any particular accountability system because the essential components relating to content, testing, passing scores, and sanctions are inherently arbitrary. The closer one gets to crafting and enforcing standards the less defensible specific program elements can appear. In the end, standards are a useful artifice. A commitment to the promise of coercive reform requires embracing a system of accountability while recognizing that such reforms will inevitably include some arbitrary and unpopular components.

Determining what students need to know, when they need to know it, and how well they need to know it is an ambiguous and value-laden exercise. Neither developmental psychologists nor psychometricians can "prove" that specified content ought to be taught at particular grade levels. Such decisions are imperfect, publicly rendered judgments about the needs and capacities of children. Because public schooling requires public officials to make these judgments and impose them statewide, these difficult questions inevitably become political ones.

In the end, there are several compromises that policymakers make as they design and then implement high-stakes accountability systems. While each can be readily

justified on practical or educational grounds, the larger point is that each marks a retreat from the transformative premise of coercive accountability.

One common compromise is to lower the stakes of the tests for students, for educators, or for both. When sanctions are weak or nonexistent, there is little incentive for teachers, low-performing students, or anyone else to worry much about test results.

A second approach is to simply make tests easier, either by lowering content standards or by adopting easier questions. This can be a politically perilous course if it is seen as signaling a public retreat from the notion of school quality. Consequently, this tack is more often taken by a Board of Education than by a legislature and is more likely to involve technical adjustments or the altering of questions than an outright reduction of the required passing score.

Third, instead of easing the test, officials can leave the tests alone but reduce the thresholds required to pass the accountability assessments. Of course, if weakening test content is difficult, the decision to formally lower the score required to pass the tests is at least equally so. Once passing scores are established, it is immensely difficult for officials to lower the bar. Consequently, the most popular way to ease the threshold is to offer lots of second chances. Giving students a number of retests or schools several years to boost their performance ensures that the law of averages will help a number of moderately low performers to clear the bar. Just as a pretty solid student might fare poorly on a given exam one time out of five, so a mediocre student may score a 70% one-fifth of the time.

If officials choose not to weaken the sanctions and find it difficult to weaken content or lower the bar, there are two other accommodations they may adopt. One is to

permit some students to side-step the required assessment by providing some form of opt-out provision. This can take the form of permitting students at the bottom to receive a "basic" diploma or completion certificate without passing the exam or that of permitting high-achieving students to substitute advanced tests for basic exams seen as interfering with "important" instruction. A second accommodation is to reduce opposition by delaying the implementation of sanctions. This permits legislators to take strong action, push the day of reckoning into the (sometimes) distant future, and mollify opponents who know that changes in the political climate or turnover among public officials may later provide a chance to modify the proposed program.

While the push for coercive accountability summons fierce opposition, there is evidence that the political dynamic may reverse if these systems can be institutionalized. Experience suggests that once high-stakes exams are in place for a sufficient period they become part of the "grammar" of schooling for educators, parents, and voters. Over time, the diffuse benefits of accountability become more evident. When high-stakes accountability is institutionalized, the tests become accepted as the unquestioned "gold standard" for measuring performance and all involved parties adjust their behavior accordingly. At that juncture, opponents of high-stakes testing find themselves in the unenviable position of attacking an established system that helps to ensure that students are learning, teachers are teaching, and that schools are serving their public purpose.

Coercive accountability only drives behavior and changes cultural norms when high-stakes regimes survive implementation. Given the challenges, proponents of high-stakes reform have stumbled upon four approaches that improve the odds of survival.

The most common approach is compromise. Reformers can reduce the size and scope of "losers" by shrinking the number of students, teachers, and schools that will be labeled inadequate by a test and/or reducing the real consequences of being deemed inadequate. This builds comfort with accountability, but it does so by lowering standards and by rendering them less significant—a price that reformers may not be willing to pay.

A second approach is to start by initially setting passing thresholds at a low level and then gradually ratcheting them up. Such an approach gives all parties a chance to gradually become acclimated to standards. It also serves to dull the effectiveness of critics, as they have little incentive to respond sharply to the minimal standards first put in place. By the time that standards are raised to more significant levels, it is difficult for critics to overcome the more accepting position they have staked out. This gradualist approach is the route that Texas followed with much success. Such an approach can also backfire, however, as proponents who settle for initially weak legislation may have trouble later raising the bar.

A third approach is to seek to make the status quo so frightening that voters will demand change and reward officials who resist efforts to weaken reforms. A number have reformers in various states have sought to employ this "*Nation at Risk*" strategy, with mixed success. The approach is alluring because it can alter the terms of the debate. However, it is difficult to whip up a widespread sense of crisis or to sustain it for any length of time, limiting the effectiveness of this approach. Moreover, the tactic is often perceived as an assault on public education and on educators, alienating centrist voters and mobilizing the opposition.

Finally, proponents can seek to make standards more palatable to educators by tamping down the leading source of opposition. One way to do this is to accelerate the turnover of teachers and administrators while ensuring that new personnel are familiarized with standards and high-stakes testing as a condition for their entry into the field. This increases the percentage of teachers trained and acculturated in an environment where high-stakes accountability is the norm. Similarly, encouraging districts to recruit more entrepreneurial administrators and to train them in the strategies of outcome-based management will help to reduce educator opposition to standards, to make the transition to standards-based schools an easier one, and to foster the ranks of public educators who are supportive of transformative accountability.

The effectiveness of high-stakes accountability rests upon the willingness of public officials to institutionalize a number of subjective and arbitrary decisions. Linking meaningful consequence to these decisions has the power to fundamentally transform schooling, especially in those schools where a reliance on educator magnanimity has failed to serve the interests of the students. Harnessing this power, however, requires standing firm on a series of difficult decisions which will visit harm and inequity upon some students and teachers. The success of such an approach depends on whether proponents can convince voters to embrace a system of accountability long enough for the accompanying benefits to take hold. In practice, the effort to enact high-stakes accountability is often met by compromising key elements of the reform. While each compromise is reasonable and softens the negative effects of coercive accountability, each also marks a retreat from the transformative promise of accountability. The question is whether proponents of high-stakes accountability are willing and able to

sustain the support required to institutionalize the proposed reforms, or whether their efforts will prove more symbolic than substantive.

Consideration of the nation's experience with accountability in the past three decades suggests an increasing willingness to accept the costs of accountability. This raises the possibility that ongoing efforts will deliver the substantive change that has often proved elusive. While testing regimes have come and gone quite rapidly, even in states with relatively stable systems, there does appear to be a growing willingness on the part of voters and public officials to stand fast in the face of inequities and concentrated costs that sank earlier accountability efforts. Public information, political efforts by pro-accountability forces, concern about school performance, a weakening attachment to local control in education, and comfort with increasingly sophisticated testing technologies all appear to be gradually shifting the center of public opinion.

A number of states appear committed to testing all students, even those with special needs, and willing to deny diplomas to thousands of graduates—whereas earlier accountability efforts faltered when they were on the verge of denying diplomas to mere hundreds. Nonetheless, legislators and policymakers have generally tiptoed up to implementation, only to declare a need for short delays or to further refine the system. Whether the ongoing efforts will prove to be small refinements or a more fundamental retreat, and whether states with more resistant political contexts will adopt such measures, are questions that will be more fully answered in the years ahead.

[1] For an illustrative example of how coercive accountability can transform school districts in positive ways, see the case studies of four high-poverty, high-achieving Texas school districts presented in Charles A. Dana Center, "Equity-driven achievement-focused school districts: A report on systemic school success in four Texas school districts serving diverse student populations" (Austin, TX: University of Texas at Austin, 2000).

[2] For a particularly effective discussion of this point, see Arthur G. Powell, Eleanor Farrar, and David K. Cohen, *The Shopping Mall High School: Winners and Losers in the Educational Marketplace* (Boston: Houghton Mifflin, 1985).

[3] Standards may potentially impact curriculum and pedagogy in very different ways. Standards presume that curricular content ought to be shaped by statewide officials rather than by individual classroom teachers. There is no similar presumption requiring the standardization of pedagogy. Rather, accountability systems will likely reward any teacher who is able to produce impressive outcomes. While some pedagogies will prove more effective than others at producing the desired results, effective teachers need not adopt any particular approach so long as they produce acceptable outcomes.

[4] For impassioned discussions of the costs linked to high-stakes accountability, see recent books by Alfie Kohn, *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools* (Portsmouth, NH: Heineman, 2000); Linda M. McNeil, *Contradictions of School Reform: Educational Costs of Standardized Testing* (Routledge, 2000); or Susan Ohanian, *One Size Fits Few: The Folly of Educational Standards* (Portsmouth, NH: Heinemann, 1999).

[5] Surveys find that adults overwhelmingly support standards and accountability in the abstract, but begin to express doubts emerge when it comes to putting high-stakes standards into practice. For instance, Public Agenda has reported that roughly eight in ten Americans believe it wrong to base grade promotion or graduation on standardized tests. This hesitance is due to the fact that two-thirds or more of Americans express concern that some students do not test well, that testing cannot measure all the skills children should learn, and that too much reliance on testing will cause teachers to focus too heavily on tested material. See Public Agenda, *Questionnaire and Fall Survey Results: National Poll of Parents of Public School Students* (New York, 2000).

[6] The fact that so many minority and low-income children are ill-served by their current schools helps to explain why leaders of these groups emerge on both sides of the high-stakes accountability discussion. While some leaders voice concerns about the inequitable impact of sanctions, others applaud the fact that educators are being forced to effectively serve students they previously overlooked.

[7] For a classic discussion, see R. Doug Arnold, *The Logic of Congressional Action* (New Haven: Yale University, 1990).

[8] National Research Council, *High Stakes: Testing for Tracking, Promotion*, and Graduation (date??), p. 45.

[9] For instance, the delegates to the 2001 annual convention of the NEA voted to oppose mandatory national testing—with or without attached consequences—and to support state legislation that permits parents to let their children skip standardized tests.

[10] The value of coercive accountability can be justified despite such concerns, as Jay Greene has argued, but it neither an easy nor alluring argument for a public official to make. See Jay Greene, "The business model," *Education Next*, vol. 2, no. 2 (2002), pp. 20-22.

[11] For the classic treatment of the resultant "school house" culture, see Dan C. Lortie, *Schoolteacher: A Sociological Study* (University of Chicago, 1975).

[12] For a useful, though ideologically colored, account of how this tension produced conservative backlash against high-stakes accountability in the early 1990s, see Fair Test *Examiner*, "Right wing attacks performance assessment" (summer 1994).

[13] See John E. Chubb and Terry M. Moe, *Politics, Markets, and America's Schools* (Brookings Institution, 1990), pp. 26-47, for a particularly nice discussion of this point in the context of education.

[14] See David B. Tyack and William Tobin, "The grammar of schooling: Why has it been so hard to change," *American Educational Research Journal*, vol. 31 (1994), pp. 453-479.

[15] Frederick M. Hess, "Reform, resistance,...retreat? The predictable politics of accountability in Virginia. In Diane Ravitch (ed.), *Brookings Papers in Educational Policy 2002* (Washington DC: Brookings Institution, 2002).

BEST COPY AVAILABLE

[16] Jeri J. Goldman, "Political and legal issues in minimum competency testing," *Educational Forum*, vol. 48 (winter 1984), pp. 207-216.
[17] Jeri J. Goldman, "Political and legal issues in minimum competency testing," *Educational Forum*, vol. 48 (winter 1984), pp. 207-216, p. 208.
[18] Chris Pipho, "Stateline", *Phi Delta Kappan* (May 1997), pp. 673-674.
[19] Darryl Paulson and Doris Ball, "Back to basics: Minimum competency testing and its impact on minorities," *Urban Education*, vol. 19, no. 1 (1984), pp. 5-15.
[20] For instance, a study of initial MCT passing rates during 1978-1981 in four states with heterogeneous populations (California, Florida, North Carolina, Virginia) showed that black students fared between 10 and 43 percentage points worse than did white students, depending on the subject and the state. Robert C. Serow, "Effects of minimum competency testing for minority students: A review of expectations and outcomes," *Urban Review*, vol. 16, no. 2 (1984), pp. 67-75.
[21] Martha M.McCarthy, "Minimum competency testing for students: Educational and legal issues," *Educational Horizons*, vol. 61 (spring 1983), pp. 103-110.
[22] Chris Pipho, "Stateline", *Phi Delta Kappan* (May 1997), pp. 673-674. p. 673.
[23] Stanley J. Vitello, "Handicapped students and competency testing," *Remedial and Special Education*, vol. 9 (September/October 1988), pp. 22-27.
[24] Robert C. Serow, "Effects of minimum competency testing for minority students: A review of expectations and outcomes," *Urban Review*, vol. 16, no. 2 (1984), pp. 67-75, p. 71.
[25] William B. Walstad, "Analzying minimal competency test performance," *Journal of Educational Research*, vol. 77 (May/June 1984), pp. 261-266.
[26] Jeri J. Goldman, "Political and legal issues in minimum competency testing," *Educational Forum*, vol. 48 (winter 1984), pp. 207-216.
[27] Thomas M. Haladyna, Susan B. Nolsen, and Nancy S. Haas, "Raising standardized achievement test scores and the origins of test score pollution," *Educational Researcher* (June/July 1991), pp. 2-7.
[28] See Public Agenda, *Questionnaire and Fall Survey Results: National Poll of Parents of Public School Students* (New York, 2000).
[29] Audrey L. Amrein and David C. Berliner, "High-stakes testing, uncertainty, and student learning," *Education Policy Analysis Archives*, vol.10, no. 18 (2002), electronic journal.
[30] Perhaps not surprisingly, legislators have been much more willing to move quickly when implementing systems of rewards for teachers or schools.
[31] Jacques Steinberg, "Student failure causes states to retool testing programs," *New York Times* (December 22, 2000).
[32] Darcia Harris Bowman, "Turf war erupts in Ariz. over delaying graduation test," *Education Week*, p. 21 (April 4, 2001).
[33] Julie Blair, "Wisconsin legislators approve revised plan for graduation exam," *Education Week*, p. 17 (October 20, 1999).
[34] Lynn Olson, "Indiana case focuses on special ed.," *Education Week*, May 31, 2000. At www.edweek.org.
[35] Gary K. Hart, "The California pupil proficiency law as viewed by its author," *Phi Delta Kappan* (May 1978), pp. 592-595.
[36] Bill Honig and Francie Alexander, 1996, "Re-writing the tests: Lessons from the California state assessment system," 143-165, pp. 144-147.
[37] Bill Honig and Francie Alexander, 1996, "Re-writing the tests: Lessons from the California state assessment system," 143-165, pp. 157-159.
[38] Michael W. Kirst and Christopher Mazzeo, "The rise, fall, and rise of state assessment in California: 1993-96," *Phi Delta Kappan*, vol. 78 (December 1996), pp. 319-323.
[39] Millicent Lawton, "States move to toughen exit exams," *Education Week*, February 21, 1996, p. 1; Gwen Stephens, "Assessment trends at the state level," *Thrust for Educational Leadership*, vol. 25 (May/June 1996), pp. 13-15.
[40] The hurried implementation led to a number of public difficulties. Most troublingly, in 1998, nearly 700 of the state's 8,500 schools got inaccurate test results, and more than 750,000 students were not included in the statewide analysis.
[41] Robert C. Johnston, "At session's close, Calif. Lawmakers OK test plan," *Education Week*, p. 17 (September 24, 1997).
[42] *New York Times* (May 20, 2001).

[43] Mike Kirst, "Swing state," *Education Next*, vol. 2, no. 2 (2002), pp. 44-49.
[44] Gary K. Hart and Nancy S. Brownell, "An era of educational accountability in California," *The Clearing House*, March/April 2001: 183-186, p. 184.
[45] Jessica L. Sandham, "Calif. board votes to pare down new graduation test," *Education Week*, p. 27 (December 13, 2000).
[46] California Teachers Association, "Bush's voucher plan, U.S. high-stakes testing idea would hurt California's and nation's public schools," (January 24, 2001). At http://www.cta.org/news/news_archive/bush_voucher.html.
[47] Jessica L. Sandham, "Calif. legislature considers postponing graduation exam." *Education Week*, p. 18 (February 28, 2001).
[48] Kathleen Kennedy Manzo, "More than half of California 9th graders flunk exit exam," Education Week, p. 19 (June 20, 2001).
[49] FairTest press release, "End the racist tests, city schools deserve the best" (May 8, 2001).
[50] Richard Rothstein, "Lessons: A rebellion is growing against required tests," *New York Times* (May 30, 2001). Kathleen Kennedy Manzo, "Protests over state testing widespread," *Education Week*, p. 1 (May 16,2001).
[51] The English Language Arts test consists of one written essay, six open response questions, and 48 multiple-choice questions and the math test of seven open-response questions, five short-answer questions, and 39 multiple-choice questions.
[52] National Research Council Committee on Appropriate Test Use, *High Stakes: Testing for Tracking, Promotion, and Graduation*. (National Academy Press: Washington DC, 1999), p. 44.
[53] The Board also voted to endorse Chairman Jim Peyser's suggestion that it commit itself to raising the required score in future years.
[54] Paul T. O'Neill, "Special education and high-stakes testing for high school graduation: An analysis of current law and policy," *Journal of Law and Education*, vol. 30, no. 2, April 2001: 185-222, p. 218.
[55] Massachusetts Department of Education. Report of 2000 Massachusetts and Local School District MCAS Results by Race/Ethnicity (Massachusetts Department of Education: Malden, MA, July 2001).
[56] Steve Rauscher, :MCAS opponents rip test as 'arbitrary' at gathering," *Boston Herald*, p. 16 (March 25, 2001).
[57] Ed Hayward, "State kicks of $500G campaign to boost public support of MCAS," *Boston Herald*, p. 12 (April 4, 2001).
[58] Lynn Olson, "States adjust high-stakes testing plans," *Education Week* (January 24, 2001).
[59] Anand Vaishnav, "Teachers, in poll, fault MCAS mandate," *Boston Globe* (April 9, 2001).
[60] Ed Hayward, "New round of tests prompts MCAS arguments," *Boston Herald*, p. 3 (May 15, 2001).
[61] Roselyn Tantraphol, "Dozens of bills target MCAS exam," *Union-News* online (June 21, 2001).
[62] Ed Hayward, "Board OKs 5 chances at passing MCAS test," *Boston Herald*, p. 3 (January 24, 2001). Michele Kurtz, "Testing testing: School accountability in Massachusetts" (Rappaport Institute for Greater Boston, Boston: 2001).
[63] Scott Greenberger, "When good students fail MCAS: Education board looks to institute an appeals process," *Boston Globe* (February 15, 2001).
[64] Ed Hayward, "Ed board official blasts MCAS on way out," *Boston Herald*, p. 8 (March 3, 2001).
[65] To be eligible, students would have to take the grade 10 MCAS at least three times; score a 216 (compared to the conventional requirement of 220) at least once; maintain a 95 percent attendance level; and participated in school-sponsored tutoring and academic support services. An appeal is to contain evidence of the student's knowledge and skills in the subject area, including teacher recommendations, the student's grades or grade point average for courses taken in the subject area, work samples and scores on other standardized tests in the subject areas. Massachusetts Department of Education Press Release, "MCAS appeals process unanimously approved" (Massachusetts Department of Education: Malden, MA, January 23, 2002).
[66] *Education Assessment Insider*, "Massachusetts approves appeals for exit test," vol. 1, no. 1 (April 2002), pp. 9-10.
[67] *Education Assessment Insider*, "State assessment news: Two-thirds of Massachusetts students fail retest," vol. 1, no. 1 (April 2002), p. 9.
[68] Anand Vaishnav, "One in four juniors still failing MCAS," *Boston Globe* (April 26, 2002).

BEST COPY AVAILABLE

[69] *Education Assessment Insider*, "State assessment news: Two-thirds of Massachusetts students fail retest," vol. 1, no. 1 (April 2002), p. 9.
[70] Jeanette DeForge, "HCC may enroll students failing MCAS," *Union-News* online (January 4, 2002).
[71] K. L. Cruse and J. S. Twing, 2000, "The history of statewide achievement in Texas," *Applied Measurement in Education*, vol. 13: 327-331.
[72] Walter Haney, "The Texas miracle in education," *Education Policy Analysis Archives,* vol. 8, no. 4 (2000), electronic journal.
[73] Walter Haney, "The Texas miracle in education," *Education Policy Analysis Archives,* vol. 8, no. 4 (2000), electronic journal.
[74] Don McAdams, *Fighting To Save Our Urban Schools... and Winning! Lessons from Houston* (Teachers College Press, New York: 2000), pp. 71-73.
[75] Susan Sclafani, "Using an aligned system to make real progress for Texas students," *Education and Urban Society* vol. 33 (2001), pp. 305-312. Texas Education Agency, Office of Accountability, "Highlights of Senate bill 7 related to accountability, assessment, public reporting and awards, accreditation, site-based decision making" (Austin, TX: Texas Education Agency, June 1993).
[76] Joan Walsh, "The truth about Texas school reform," *Salon.com* (November 1, 1999). At http://www.salon.com/news/feature/1999/11/01/elpaso/
[77] Paul T. O'Neill, "Special education and high-stakes testing for high school graduation: An analysis of current law and policy," *Journal of Law and Education*, vol. 30, no. 2 (April 2001), pp. 185-222, p. 212.
[78] *Education Assessment Insider*, "State profile: Texas raises the bar amid high-stakes testing controversy," vol. 1, no. 1 (April 2002), p. 8.
[79] Craig D. Jerald, "Real results, remaining challenges: The story of Texas education reform" (Washington DC: Business Roundtable, 2001), p. 10.
[80] Walter Haney, "The Texas miracle in education," *Education Policy Analysis Archives,* vol. 8, no. 4 (2000), electronic journal. Debra Viadero, 2000, "Testing systems in Texas yet to get final grade," *Education Week*, May 31, 20-21. For an extensive attack on the merits of the Texas accountability system, see Linda McNeil, *Contradictions of School Reform: Educational Costs of Standardized Testing* (Routledge, 2000).
[81] Fair Test *Examiner*, "More errors plague testing" (summer 2001).
[82] The Education Trust, "Ticket to nowhere," (Washington DC, 1999). At www.edtrust.org/main/reports.asp.
[83] *Education Assessment Insider*, "State profile: Texas raises the bar amid high-stakes testing controversy," *Education Assessment Insider*, vol. 1, no. 1 (April 2002), p. 8.
[84] Public Education Integrity Task Force, "Public school accountability system: Texas assessment of academic skills (TAAS)." Chapter 3 in *A Report to the Comptroller* (Texas Comptroller's Office: Austin, TX, 2001). At http://www.window.state.tx.us/tspr/peitf/chapt3a.htm.
[85] Maggie Locke, "Fairfax postpones minimum competency tests in math, reading," *Washington Post*, Virginia Weekly 16 (October 12, 1978).
[86] Gary Robertson, "Competence tests scored," *Richmond Times-Dispatch*, B1 (August 7, 1979) and Lawrence Feinberg, "18% of Virginia 10th graders fail reading, math tests," *Washington Post*, C1 (January 17, 1979).
[87] These results mirrored the results of every previous state to adopt minimum competency testing. In every case, less than 1% of students failed the test when high school graduation was at stake. Lawrence Feinberg, "High school competency tests viewed as too easy," *Washington Post*, A1 (December 8, 1981).
[88] Peter Baker, "Test scores are high but show some stubborn gaps," *Washington Post*, Virginia Weekly 1 (July 5, 1990).
[89] See Frederick M. Hess and David L. Leal, "Republican party growth in southern legislative elections," *Virginia Newsletter*, vol. 75, no. 4 (1999), pp. 1-4.
[90] The social studies standards encompassed the disciplines of civics, geography, economics, and history, but it was history that was source of most conflict.
[91] Teresa Lemons Coleman, "Social studies: 'A long shot'," *Richmond Times-Dispatch*, A1 (June 18, 1995).
[92] Gary Robertson, "Testing, counselors get more study," *Richmond Times-Dispatch*, B7 (November 17, 1995).

BEST COPY AVAILABLE

[93] The executive director of the Virginia chapter of the NAACP, Salim Khalfani, said, "Our concerns are that [black students are] going to be steered and tracked into that kind of thing. It flies in the face of their SOL effort if our young people aren't getting what they need to meet the work force of the 21st century." Pamela Stallsmith, "'Basic diploma' may be revised," *Richmond Times-Dispatch*, C1 (July 23, 2000).

[94] For more extended discussion, see Frederick M. Hess and Frederick J. Brigham, "How federal special education policy affects schooling in Virginia," in Chester E. Finn, Jr., Andrew J. Rotherham, and Charles R. Hokanson, Jr. (eds.), *Rethinking Special Education for a New Century* (Washington DC: Thomas B. Fordham Foundation and Progressive Policy Institute, 2001), pp. 161-181.

[95] Alice Digilio, "New rating plan for Va. schools viewed as overly bureaucratic," *Washington Post*, B5 (July 14, 1990).

[96] Pamela Stallsmith, "'Proposal to modify SOL tests' role advances," *Richmond Times-Dispatch*, A8 (January 27, 2001).

[97] See Lawrence H. Cross, "Are Virginia's public schools failing? Assessing the assessments," *Virginia Issues and Answers*, vol. 6, no. 1 (1999), pp. 2-6.

[98] Rhea R. Borja, "Statewide test scores rise," *Richmond Times-Dispatch*, A1 (September 8, 2000). Pamela Stallsmith, "Lawmakers Seeking SOL Changes," *Richmond Times-Dispatch*, C1 (January 24, 2001). Roxanne Grossman, a spokeswoman for Parents Across Virginia United to Reform SOLs said in early 2001, "We're not opposed to including standardized testing scores in the criteria [for determining graduation and accreditation], but what needs to be gotten at is that the criterion of the test scores should not be given the most weight."

[99] Jay Matthews, Victoria Benning, Washington Post staff writers, "Va. Voters Negative on SOLs," *Washington Post*, B1 (September 11, 2000).

[100] Jeff E. Schapiro, "Survey: SOL tests lose public support," *Richmond Times-Dispatch*, A1 (October 6, 2000).

[101] When AFT President Sandra Feldman denounced testing proponents at the union's 2001 national convention for failing to provide teachers with the necessary texts and teaching materials, she framed her remarks as a reasonable caution while implicitly advocating an open-ended suspension of coercive accountability measures. The irony is that Feldman's AFT has officially endorsed the use of high-stakes testing for graduation and promotion decisions.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title:	I Say 'Retiring', You Say Retreating ...
Author(s):	Frederick M. Hess
Corporate Source: PEPG Conference	Publication Date: Presented June 10-11, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproductic and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.

If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminc document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC empl system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by librari service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature:	Printed Name/Position/Title:	
	Frederick M. Hess, Resident Schola	
Organization/Address:	Telephone: 202-828-6030	Fax: 202-862-7178
American Enterprise Institute 1150 17th St., NW Washington DC 20036	E-mail Address: rhess@aei.org	Date: 10-31-02

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

Document Acquisitions Department
ERIC Clearinghouse on Educational Management
5207 University of Oregon
Eugene, Oregon 97403-5207